

POLICY RESEARCH WORKING PAPER

7054

# Water Quality, Brawn, and Education

## The Rural Drinking Water Program in China

*Lixin Colin Xu*

*Jing Zhang*



**WORLD BANK GROUP**

Development Research Group

Finance and Private Sector Development Team

October 2014

## Abstract

Although previous research has demonstrated the health benefits of water treatment programs, relatively little is known about the effect of water treatment on education. This paper examines the educational benefits to rural youth in China of a major drinking water treatment program started in the 1980s, perhaps the largest of such programs in the world. By employing a cross-sectional data set (constructed from a longitudinal data set covering two decades) with more than 4,700 individuals between 18 and 25 years old, the analysis finds that this health program has improved the individuals' education substantially, increasing the grades of education completed by 1.08 years. The qualitative results hold when the analysis controls for local educational policies and resources, village dummies, and distance of villages to

schools, and by instrumenting the water treatment dummy with villages' topographic features, among others. Moreover, three findings render support to the brawn theory of gender division of labor: girls benefit much more from water treatment than boys in schooling attainment; youth with an older brother benefit more than youth with an older sister; and boys gain more body mass than girls do from having access to treated water. The program can account for the gender gap in educational attainment in rural China in the sample period. Young people that had access to treated plant water in early childhood (0–2 years of age) experienced significantly higher gains in education than those who were exposed to treated water after early childhood. The estimates suggest that this program is highly cost-effective.

---

This paper is a product of the Finance and Private Sector Development Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at [lxu1@worldbank.org](mailto:lxu1@worldbank.org).

*The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.*

# **Water Quality, Brawn, and Education: The Rural Drinking Water Program in China<sup>1</sup>**

Lixin Colin Xu

World Bank

Jing Zhang

Renmin University of China

**Keywords:** Water treatment, education, brawn, gender, fiscal programs.

**JEL codes:** I0, I1, I20, J16, O10.

---

<sup>1</sup> We thank Hanan Jacoby, Gary Libecap, Mingxing Liu, Liping Lu, Jintao Xu, Yang Yao, Junjian Yi and other participants of seminars at the Walter H. Shorenstein Asia-Pacific Research Center at Stanford University, the Bren School of UC Santa Barbara, Peking University, Shanghai Jiaotong University, Renmin University of China and Central University of Finance and Economics for their useful comments. We are especially grateful for substantial and constructive comments by Sebastian Galiani,

Corresponding authors: Jing Zhang, China Financial Policy Research Center, School of Finance, Renmin University of China, 59 Zhongguancun Street, Beijing 100872, China, [zhangjingecon@gmail.com](mailto:zhangjingecon@gmail.com); L. Colin Xu, mail stop MC 3-307, World Bank, 1818 H Street, N.W. Washington, DC 20433, [lixin.colin.xu@gmail.com](mailto:lixin.colin.xu@gmail.com).

## 1 Introduction

An unfortunate consequence of industrialization in developing countries is the contamination of drinking water, as chemical impurities associated with untreated industrial waste and excess use of agricultural fertilizers and pesticides have become major water pollutants in more and more countries. About 1.5 million episodes of skin lesions and one million of skeletal fluorosis are related to the poor quality of drinking water (World Health Organization, 2004). There is no doubt that industrial contamination of drinking water has contributed to the drinking water crisis in the world, with still 884 million people relying on unsafe drinking water in the world as of 2008 (World Health Organization, 2011). Despite significant welfare consequences associated with chemical impurities in drinking water, studies on the effect of treated water remain limited. Previous research has found significant effect of treated water on health (Zhang, 2012). In this paper, instead of focusing on the health consequence, we examine its longer-term benefits on youth's final educational attainment following the rollout of the rural drinking water program, the same program as in Zhang (2012). In light of recent literature that emphasizes differential impacts of health intervention on gender due to gender-specific impact of health on brawn (Pitt et al. 2012), and the critical importance of early childhood health and nutrition (Almond and Currie, 2011; Chnha et al., 2006; Heckman, 2008), we also explore how the effects of treated water differ by gender and by age of first exposure.

The key challenge for studying the educational impacts of health in general and water quality improvement in particular is to uncover the causal effects. Many studies rely on observational data to explore the correlations between health and education at the individual level without having access to exogenous variation based on the introduction of a specific health intervention (e.g., Dean, 1986; Glewwe et al., 1999; Alderman et al., 2001; Glewwe et al., 2001; Alderman et al., 2006). However, without an explicit (potentially exogenous) health intervention behind the variation in health, the educational effects based on the regressions of education on health are likely to be contaminated by biases from selection on unobservables such as personal and family traits that affect both health and education. Other studies employ randomization and experimental methods to create exogenous health variation, which helps to avoid omitted variable bias more convincingly (e.g. Pollitt et al., 1993; Martorell et al., 1995; Nokes et al.,

1998; Dickson et al., 2000; Miguel and Kremer, 2004; Vermeersch and Kremer, 2005; Bobonis et al., 2006). Nevertheless, in experimental studies one observes only short-term impacts due to a limited period. Yet short-term positive impacts of health on education (such as school attendance) may not necessarily imply improvement of individuals' final education. Indeed, since health is also positively correlated with individuals' labor market outcomes such as wage rates (Thomas and Strauss, 1997; Thomas et al., 2006), labor supplies (Strauss and Thomas, 1998; Thirumurthy et al., 2008), self-employment profits (Singh et al., 1986; Strauss, 1986), and agricultural productivities (Audibert, 1986; Strauss, 1986), a youth may choose to join labor markets rather than to stay in school. So we could find short-term gains in employment and income but long-term losses in education. Indeed, Pitt et al. (2012) show that this scenario is more likely to happen on male youth since their comparative advantages in brawn-intensive work lead to a stronger pull to the labor market for male than for female when their health improves. Thus, even if health interventions from randomized experiments improve some educational outcomes instantly, such as class attendances and academic grades, their impacts on individuals' final education remain uncertain.

To obtain a better understanding about the relationship between health and education, it is thus helpful to explore the *final* educational effects of health-oriented government programs. Two such studies in the U.S. find some educational benefits of health programs. Bleakley (2007) finds that US hookworm eradication in the 1910s improved school enrollments of children between 8 and 16 years old, but showed a significant increase only in quality of education instead of quantity of education (years of schoolings) for those exposed cohorts in the long term. Quynh (2010) finds that the introduction of iodized salt in U.S. in 1924 improved boys' schooling and raised men's income. Our analysis follows this line of research by studying the educational effects of the rural drinking water program in China.

One of the most ambitious programs for improving water quality for poor people, this rural water program, starting from the early 1980s, had incurred a cost more than \$8.8 billion U.S. dollars by 2002 (Meng et al. 2004) and had covered around 300 million people by 2008 (Center for Health Statistics and Information, Ministry of Health of the People's Republic of China, 2009). This program aims to build up water plants and pipelines to provide rural residents with treated drinking water. A key component of the water treatment plants, besides improving

water access, is to eliminate the chemical contaminants and microorganisms by using clean water technology and equipment. Few studies so far have been conducted to evaluate this important program.<sup>2</sup> In this paper, by employing the longitudinal data of the China Health and Nutrition Survey (CHNS) from 1989 to 2011, we study the long-term educational benefits of the improvements in drinking water *quality*.

Our investigations suggest that young people in villages with access to plant water have better education than those without such access: the grade of education completed among youth increases by 1.08 years, and the likelihood of graduating from a middle and high school increase by 19.8 and 89.6 percent, respectively. The results are obtained after controlling for county-year dummies (and therefore local educational policies and resources), household characteristics, and village characteristics including the distances to schools. These results remain qualitatively robust after dealing with the endogeneity of the water treatment program that relies on topographic features of villages, controlling for village fixed effects, considering local labor market conditions, controlling for water access (i.e., whether a household has access to water in its premise), and estimating the heterogeneous effects by income groups and by gender. The main channels through which plant water benefits youth's education are improved health of the youth themselves as well as increased household income resulted from improved health of other household members and early entry into brawn-typed jobs of their elder brothers if any, but not time saving due to a better *access* to water.

We also find strong support for the brawn-based theory of gender division of labor in Pitt et al. (2012). In particular, boys gained more body mass than girls from using plant water; girls benefit much more from water treatment than boys in terms of schooling attainment, and youth with an older brother benefit more than youth with an older sister. The program benefited girls so much more than boys such that it completely wipes away the gender gap in educational attainment in treated areas of rural China.

Young people that had access to treated plant water in early childhood (i.e., 0-2 years of age) are found to experience significantly higher gains in education than those who were exposed to treated water after early childhood, consistent with recent literature on the critical importance

---

<sup>2</sup> An exception is Zhang (2012), which examines the benefits on health of this program.

of early childhood for human capital investment (Almond, 2006; Cunha et al. 2006; Heckman 2008; Maccini and Yang, 2009; Cunha and Heckman 2010; Almond and Currie 2011). Our estimates suggest that this program is highly cost-effective.

Our paper contributes to four strands of literature. The first is the literature of the effect of safe drinking water programs. While the literature on water programs is not small (see Esrey et al., 1991; Tonglet et al., 1992; Jalan and Ravallion, 2003; Fewtrell et al. 2005; Galiani et al., 2005; Clarke et al., 2009; Maimaitwe and Siebert, 2009; Gamper-Rabindran et al., 2010; Kremer et al., 2011), there have been few studies that rigorously examine the effect of *water treatment and water quality*, especially on final educational attainment. Kosec (2014) uses child-level data from 39 African countries and finds that private sector participation in the piped water sector decreases diarrhea among urban under-five children, and is associated with a 8 percentage point increase in school attendance of 7-17 year olds.<sup>3</sup> However, the water treatment program in Kosec (2014) concerns access to piped water. It does not deal with change in water quality, and nor does it examine the long-term effect on final schooling attainment. We differ from this literature in focusing on water treatment program of which changes in water quality is a key component, examining the long-term effects on children's final educational attainment, along with how the effect differs by gender and exposure at different periods of the early life cycle. The second is the literature of human capital investment and the gender division of labor pioneered by Pitt et al. (2012). We add to this literature by providing a dramatic example of how a large-scale health intervention ends up improving rural girls' education much more than rural boys' education through the brawn channel. Indeed, the relative female gains over male is so big that the rural gender gap in education is completely eradicated—with all this being accomplished within a short span of two decades. The third contribution is to the literature of the long term effects of early childhood conditions (see Maccini and Yang, 2009; Cunha and Heckman, 2010; Almond and Currie, 2011). We show that exposure to treated water at very young age (i.e., 0-2 years of age) had much more pronounced effect on a person's final educational attainment than exposure at later life stages.<sup>4</sup> Finally, there have been few studies that examine the cost effectiveness of

---

<sup>3</sup> Galiani et al. (2005) also find strong benefits on child mortality of private sector provision of water, but they do not examine the impact on education.

<sup>4</sup> Maccini and Yang (2009) find that Indonesian women experienced long-term gains in education (and other socioeconomic outcomes) when they had favorable weather shocks at their birth years. No such effects exist for

fiscal programs in China. Given the magnitude of China's fiscal pie—its fiscal spending is about 14 trillion yuan (i.e., 2.3 trillion U.S. dollars) in 2013—more evaluations of specific programs are called for, and ours here is a simple example of high cost-effectiveness.

## 2 The Rural Drinking Water Program in China

Before the 1980s, the rural residents in China largely relied on untreated water from wells, rivers, and lakes. Indeed, more than 70 percent of rural households in the CHNS data drank untreated water even in 1989. The sanitation environment was also poor as human and livestock waste was disposed freely around dwellings within villages. These unsanitary practices caused endemics of water-related diseases. It should be noted that microorganisms, the major drinking water pollutants in many other developing countries, have less adverse consequences in China due to the Chinese tradition of drinking boiled water and eating cooked food (Braudel, 1982; Zhang et al., 2009). In contrast, another key water pollutant, chemical impurities such as toxic metals and inorganic and organic compounds, causes as much health damage in China as in elsewhere. Partly the result of geography as natural soil and rock contain high levels of chemicals, chemicals in drinking water are increasingly caused by the fast industrialization process coupled with weak government regulations in China (Cai et al. 2011). Vast discharges of industrial waste and excessive use of fertilizer and pesticides have led to widespread and severe water pollution, causing the spread of diseases and even deaths. In 2006 a total of 1,115 counties and about 81.6 million people are at the risk of fluorosis via drinking water.<sup>5</sup> Around 66,000 people likely die from water pollution in rural China every year (World Bank, 2007).

After largely completing the urban public water system by the 1980s, the Chinese government started the rural drinking water treatment program, and spent a significant amount of resources trying to improve the *quality* of drinking water in rural China. The ultimate goal of the program has been to build up water treatment plants where clean water technology and equipment can be installed to eradicate chemical pollutants and microorganisms and where

---

Indonesia men. They interpreted this as evidence of gender discrimination—parents allocated limited food to boys rather than girls in time of food scarcity.

<sup>5</sup> Based on Chinese National Health Statistics (2007).



related government bodies can monitor water quality precisely and regularly.<sup>6</sup> The drinking water from those water treatment plants is supposed to meet a variety of standards, including general chemical, toxicological, bacteriological and radiative indices stipulated by Sanitary Standard for Drinking Water Quality (Ministry of Health, 2007). To avoid water contamination during transport, pipeline systems are also constructed to deliver water directly from plants to households.

The implementation of the rural drinking water program has been rather decentralized. While the central government stipulates general guidelines about the locations of water plants, safe drinking water standards and monitoring, local governments are responsible for program implementation. The financing of the program is shared among the central and local governments, villages, households, and international organizations. The specific split varies greatly across regions. The western areas, for instance, rely mostly on outside donors, which account for as much as over 50 percent of the total fund. In contrast, local governments in rich areas self-finance the program or partly rely on private capital. Overall, total spending for this rural drinking water program was about 8.8 billion U.S. dollars from 1981 to 2002, and the cost for the program was approximately 30 U.S. dollars per capita (Meng et al. 2004). The rural drinking water program is far from being complete, however. Even in 2008, only 41.9 percent of rural residents had access to plant water, and about 300 million rural people were still using untreated drinking water (Center for Health Statistics and Information, Ministry of Health of the People's Republic of China, 2009).

Given the importance of drinking water quality and the sheer scale of investment in the rural drinking water treatment program in China, evaluations of its benefits are clearly important. So far, we are only aware of two studies that attempt to evaluate the program. They find that the quality of plant water for the sample villages after the implementation of this program is indeed better than that of untreated water (Zhang et al. 2009), and that this program has improved health of adults and children (Zhang 2012). Differing from the above studies, in this paper we focus on

---

<sup>6</sup> Because of the diversity of natural conditions, deep well pumps and rainwater harvesting systems have been introduced as temporary substitutes in some areas.

the effects on youth education, longer-term impacts of rural water treatments, and we pay particular attention to gender-specific and exposure-time-specific effects.<sup>7</sup>

### 3 Sample, Variables, and Estimation Strategy

We primarily rely on the China Health and Nutrition Survey (CHNS), which includes nine waves of survey so far in 1989, 1991, 1993, 1997, 2000, 2004, 2006, 2009 and 2011, respectively. CHNS covers nine provinces including Guangxi, Guizhou, Heilongjiang, Henan, Hubei, Hunan, Jiangsu, Liaoning, and Shandong. The subsample in each of the provinces was selected based on a multistage, random cluster sampling process. Since the rural drinking water program has only been implemented in rural areas, we only use the rural sample of CHNS.

Our sample consists of young people between the age 18 and 25, a period in which the school-to-work transition for rural youth is largely complete. The starting age of 18 was chosen because China's Compulsory Education Law (CEL), which took effect in 1986, mandates that children must enroll in school by age six (in some areas it can go up to age seven), and thus the vast majority of rural youth presumably would have finished their high school had they chosen so at age 18. Figure 1 plots the proportions of individuals who are currently working and in schools for each age group from 12 to 40 in our sample between 2004 and 2011.<sup>8</sup> Over 80 percent of children at age 15 stay in school, but the majority of people over 25 (87 percent) work.<sup>9</sup> In our sample around 65.8 percent of the individuals in this age group graduate from a middle school, but only 16.4 percent of them actually graduate from a high school, and less than 5 percent continue their study after that. The vast majority of the rural youth clearly do not pursue more

---

<sup>7</sup> Our paper is also related to Maimaitwe and Siebert (2009), who use the CHNS data to test the hypothesis that poor *access* to water hurts girls' education due to unmet female hygiene needs after menarche. The major distinction between their study and this paper lies in the discussion on the roles of water *access* versus water *quality*. While their focus is on water access (tap water in the house or courtyard), ours is instead on water quality. In the robustness check, we control for households' water access status in the regressions, and our baseline treatment effects remain intact. Not dealing with the improvement in water quality, the central element of the rural drinking water program, their paper does not evaluate this specific program's impacts. Moreover, we pay particular attention to gender-specific and exposure-time-specific effects which do not concern their paper.

<sup>8</sup> CHNS only recorded work status of individuals aged above 16 before 2004.

<sup>9</sup> According to the CHNS survey, people not currently working includes housewives, the disabled, students, retirees and those on job hunt.

In the waves before 2000, the work status is recorded for children whose are 16 or older. This question is also asked for children not in schools in the waves after 2004. To be consistent across waves, we only plot the work status of children over 16

schooling beyond middle school. In this age group, young rural residents thus almost always have already finished their schooling.<sup>10</sup>

### **The Treatment Variable**

We code a village in a particular year as being covered by the water improvement program when either of the two conditions hold: (1) over 80 percent of village households have a water plant as their water source in the first year; (2) less than 80 percent of village households enjoy a water plant in the first wave, but plant coverage rises by more than 20 percentage points per year since the last wave.<sup>11</sup> Once having access, a village is assumed to always have access in subsequent years. Whether a household has access to water plant is based on a survey question that is answered by the household about its water source, which includes water plants, wells, springs, and rivers. Figure 2 shows the trend of the coverage of plant water in our sample. Starting at the nadir of 20 percent of the sample having plant water in 1989, the ratio rose to 47 percent in 2011.<sup>12</sup>

### **Specification and Empirical Strategy**

Before we proceed to empirical specification, it is useful to be clear about our estimation sample. Since we are interested in the final schooling attainment of rural youth, for each person between 18 and 25, we only keep *one observation per person*. The observation present the final year for all the years that she/he is in the sample so that we know her or his final schooling attainment. Our sample is thus of the nature of a cross-sectional sample. However, since the final year differs for each individual, each observation could be in different year, and we shall use  $t$  to indicate this time dimension. Our final sample consists of 4,729 observations.

---

<sup>10</sup> As discussed later, our main results remain intact when restricting the sample to be young people of 19-25, or 20-25, or 21-25 years of age.

<sup>11</sup> For example, if over 40 percent of households in a village report that their water sources switched to plants from 1989 to 1991, then Water Plant is set to 1.

<sup>12</sup> The ratio decreases slightly in 2004 as compared to 1997, from 30.7 to 29.1 percent, due to a slight change in the survey areas. Heilongjiang province was initially surveyed as a substitute for Liaoning in 1997, and both provinces were included since 2000. In 2004 CHNS expanded the number of their surveyed villages.

To gauge the effects of the rural drinking water program on youth education, we estimate the following equation:

$$Y_{ivt} = X_{ivt}\beta + T_{vt}\gamma + w_{ct} + \varepsilon_{ivt} \quad (1)$$

Here  $i$  indicates an individual,  $v$  a village,  $c$  a county, and  $t$  a year.  $Y_{ivt}$  is a person's educational status such as grades of education completed and the dummy variables of being a graduate of a middle school and of being a graduate of a high school.<sup>13</sup> While strict enforcement of the Compulsory Education Law should lead to a 100 percent graduation rate from middle schools, its enforcement had been weak until recently. The dummy variable of being a graduate of a middle school is therefore still a meaningful measure of the youth's educational achievement.

$X_{ivt}$  refers to the characteristics of individuals, households, and villages, including the individual's age, gender, and the relationship to the household head.<sup>14</sup> We also control for household income in the first wave rather than the income in the current wave. Since water treatments also benefit adult health which further increases household income and youth education, the inclusion of the current household income as an explanatory variable may lead to an underestimation of the program's impacts on the youth--current household income absorbs part of the benefits of a water plant on the households.

$T_{vt}$  is the treatment variable, Water Plant. Note that Water Plant is one as long as the village in which he/she resides had access to treated plant water in some part of his/her sample years.<sup>15</sup> Thus in the base regressions, we do not allow the treatment effect to depend on the earliest age at which the youth had exposure to plant water; later we shall distinguish exposure-year effect.  $w_{ct}$  represents the county-year fixed effects.  $\varepsilon_{ivt}$  is the error term. A consistent estimate requires the following condition to be satisfied:

---

<sup>13</sup> These two dummy variables are constructed based on the grades of education completed. If it is greater or equal to 9, the dummy variable of being a graduate of a middle school is set to one, and if it is over 12, then we consider the individuals as a graduate from a high school. We should mention that the CHNS questionnaire uses different terms for the middle and high schools: i.e. the lower middle school and the upper middle school.

<sup>14</sup> The distances to schools were not surveyed in the first wave. Therefore, we use the information in 1991 as the proxies for the information in 1989.

<sup>15</sup> Here the subscript  $t$  merely indicates that the observation year for each individual could differ. Our treatment variable here is essentially a cross-sectional measure.

$$E(T_{vt}\varepsilon_{ivt}|X_{ivt},w_{ct}) = 0 \quad (2)$$

Our conjecture is that the rural drinking water program likely benefits education in rural China, especially that of girls. Drinking water of better quality improves young people's health, which allows them to improve their educational competency through the reduction in absenteeism and the improvement in mental focus and energy levels (Alderman et al., 2001).<sup>16</sup> Moreover, access to water of better quality can also improve the health of other household members and therefore their income, which may reinforce the educational benefits of the children due to the income effect. While this health intervention on drinking water likely benefits education of a youth, there is no guarantee. Indeed, since health is also positively correlated with individuals' labor market outcomes such as wage rates (Thomas and Strauss, 1997; Thomas et al., 2006), labor supplies (Strauss and Thomas, 1998; Thirumurthy et al., 2008), self-employment profits (Singh et al., 1986; Strauss, 1986), and agricultural productivities (Audibert, 1986; Strauss, 1986), a youth may choose to join labor markets rather than to stay in schools. So the program could result in gains in employment and income yet losses in education. Moreover, this tendency for favoring work instead of schooling could differ by gender. Since a male has more brawn and thus comparative advantage in brawn-intensive (and unskilled) work (Pitt et al. 2012), the market pull for the unskilled male would be stronger. The male, benefiting from better health by adding brawn, may therefore favor brawn-intensive work and forgo schooling to a greater extent than the female. This consideration suggests that the relative educational benefits of water treatments should be smaller for the male. Furthermore, when a rural young has an elder brother (relative to the case of having an elder sister), the elder brother may gain more brawn from the water treatment, may transit from school to work earlier, and thus raise household income more. This additional income effect associated with having an elder brother may benefit the education of the young to a greater extent than the case of having an elder sister. In our empirical work we shall test the two implications directly. We shall also provide evidence that water treatment indeed improves brawn more for the male than for the female.

An empirical issue is what level of regional fixed effects should be controlled for. This depends on what is omitted in the residual that determines the educational outcomes. We believe the key omitted determinants of local educational attainment are related to the supply side of

---

<sup>16</sup> See Bleakley (2010) for a summary of evidence on how health affects human capital and development.

education—the demand for education could be controlled for more easily by including household and village characteristics. Middle schools in rural China are usually administered at the township level. The financing of the schools was provided by township governments which collected fees mainly from rural parents before 2001, and by county governments which have relied on intergovernmental transfers ever since (Liu et al. 2009). High schools have been managed by county governments. The key determinants of local educational attainment are thus county-level educational finance and local preferences for education, both of which are plausibly fixed at the county or county-year level. We therefore choose to use the full set of county-year fixed effects instead of the village fixed effects.

Another reason not to use village fixed effects in the outcome equation is that 74 percent of villages experienced no changes in the treatment status in the sample period. During the 21 years of CHNS coverage, 89 of the 174 villages had never implemented the water program, 40 of them had plant water in all the waves, and only 45 villages changed their treatment status (see Figure 3). The employment of village fixed effects implies that the estimation only exploits the variations in those 45 villages in relatively shorter periods of time.<sup>17</sup> More importantly, we believe a key source of variation is the within-county inter-village comparisons. In light of our findings presented later that the most pronounced effect of water treatment is observed for those youth who had their first exposure to water treatment between age zero and two, relying exclusively on the treatment-status-changing villages has an unfortunate consequence: we have to rely more on youth who got exposed to plant water at age older than 2 than in the case of keeping always-treated villages in the sample.<sup>18</sup> Indeed, having the always-treated villages in the sample confers important benefits—in such villages the youth are likely to be covered by this drinking water program for a few years before they first appear in our sample. Thus, many of the youth in these villages may start to use plant water before age two, and their inclusion would allow us to better capture the returns to treatment at early childhood. With this consideration in mind, our main empirical strategy is to hold time-variant educational financing and county policies constant, to control for key village characteristics and key household characteristics, and

---

<sup>17</sup> Since the treatment happens in the middle of the sample period, the number of post-treatment years is more limited than the sample of “always treated”. This poses another challenge for identifying the long-term effect of the treatment on education.

<sup>18</sup> In particular, for data of wave 1997 and after, those individuals who got exposed to treated water at age zero to two would be excluded from the treated sample.

then to attribute inter-village differences in educational outcomes to inter-village differences in water treatment status. We shall later show that the treatment effects remain robust, though quantitatively smaller, when controlling for more village characteristics including the village fixed effect.

#### **4 Estimation Results**

Table 1 presents the basic descriptive statistics for our sample individuals aged 18 to 25. The average grades of education completed in the sample is 8.7 years, slightly less than nine years as required by the Compulsory Education Law. About 65.8 percent of individuals finish middle schools, and only 16.4 percent of individuals graduate from high schools. Roughly 74.6 percent of the sample appears as a child of a survey household. The distances to schools are substantial. While the average distance to a primary school is only 0.57 km, that to a middle school is 2.03 km, and that to a high school is 8.49 km.

##### **Baseline Results**

Table 2 presents the county-year fixed effect (FE) regression results for each outcome variable. All standard errors are clustered at the village level to allow the unobservable at the individual level to be correlated within a village. Holding constant county-year fixed effects, along with household and village characteristics (e.g., age, gender, generational status, household size, initial household income, and distances to various types of schools), having access to treated plant water is associated with an increase in schooling by 1.08 years. The linear probability model (LPM) estimates imply that access to treated plant water is associated with increases in the probabilities of graduating from a middle and high school by 13 and 14.7 percentage points, respectively.<sup>19</sup> Given that the means of these two dependent variables are 65.8 and 16.4 percent,

---

<sup>19</sup> We have checked the robustness of the results with alternative functional form. The marginal effects based on the Probit model are very similar to what we find in Table 2, both qualitatively and quantitatively. Thus, in the rest of this paper, we mainly rely on the LPMs.

the regression results imply 19.8 and 89.6 percent increases in the probabilities of graduating from a middle school and a high school after implementing the drinking water program.<sup>20</sup>

The coefficients of other covariates also make sense. Conditioning on Water Plant, females have some disadvantages in graduation from middle schools, and are otherwise non-disadvantaged. Thus, females' unconditional disadvantage is completely explained by the water treatment program. Individuals with parents or grandparents in the household tend to have more education. The household size and the number of children in the household are negatively associated with youth education in general. Young people in wealthier households have more schooling and are more likely to finish a middle school than their poorer peers. The distance to a high school is significantly associated with a lower schooling level. The estimation results suggest that, when the nearest high school is one standard deviation (i.e., 11.6 kilometers) farther, the grades completed would on average decrease by 0.29 years, and the probabilities of graduating from a middle school and a high school would drop by 3.5 and 2.3 percentage points, respectively.<sup>21</sup> Slightly smaller effects are found for the distance to a middle school.

How large is our estimate of the effect of access to plant water on schooling attainment compared with those in the literature on various types of health facilitation programs? Miguel and Kremer (2004) find that a less than two-year treatment with deworming drugs in Kenya led to a 0.14 year increase in schooling. Field et al. (2009) find that iodine supplements in utero in Tanzania increased schooling attainment by 0.35-0.56 years.<sup>22</sup> The treatment effect as suggested by the baseline regression—a gain of 1.08 years of schooling—thus represents a strong effect compared with what are found about other health programs in the literature. One thing to keep in mind is that the estimated return of this program does not happen instantaneously, but rather after more than 9.1 years of exposure to this program on average.<sup>23</sup>

---

<sup>20</sup> While the 89.6 percent increase in the probability of graduating from a high school seems to be large, it is not surprising once one considers that the mean of the outcome variable is low at 16.4 percent.

<sup>21</sup> Why does the distance to the nearest high school affect the probability of graduating from a middle school? The reason is that the value of middle school drops when the chance to attend high school declines with a longer physical distance. The force of option value is at play here.

<sup>22</sup> In Field et al. (2009), the increase in education is not due to improved health status of those babies but improved cognitive skills.

<sup>23</sup> For always treated village in the first wave when they were included in the survey, we code the years of exposure as one. This likely understates the value of years of exposure for this group. The mean of years of exposure to plant water is 9.01 years for the our sample.



While here we define Water Plant partly based on the changes in plant water coverage in a village, we do consider alternative definitions. The results are quite robust. In Table A1, we construct a variety of treatment variables by employing different cutoffs of coverage changes or by relying on cutoffs of the coverage directly. The estimates of the treatment effects remain similar across these definitions. For example, the estimate of the gain in the grades of education completed is in a narrow band from one to 1.2 years across all of these definitions.

Another legitimate concern is about our sample composition. We have used the observation of the last year between age 18 and 25 for each individual for his/her final schooling attainment. Age 18 is supposed to be the year of high school graduation, and the vast majority of rural youth would have finished their schooling. However, sometimes rural children delay their schooling, which would make graduation from high school slightly later than age 18. We thus repeat our baseline regressions, but the final sample would consist of the observation of the last year between the age 19 (or 20, 21) to 25 for each individual. The new restrictions reduce the sample somewhat (to 4572, 4402, 4207), and the estimate of the effect of Water Plant on grades of education completed become 1.09, 1.12, and 1.15, respectively, all statistically significant at the 1 percent level.<sup>24</sup>

### **Cost Effectiveness**

While the effect of the rural drinking water program is large, it remains unclear whether the program is cost effective. Yet some notion of costs and benefits of such program is important. After all, China's fiscal expenditures have been increasing dramatically over time: the total government expenditure was 706.1 billion Chinese yuan in 1989, and it becomes 14 times larger in 2011 (10,924.8 billion) (China Statistical Yearbook, 2012).<sup>25</sup> With such a grand scale of government spending, it is important to understand whether the fiscal resources are allocated cost-effectively, and this requires knowledge about the returns to various social programs funded by fiscal expenditures. To evaluate the current program and to help guide future fiscal allocation, we thus conduct a back-of-the-envelope analysis of the costs effectiveness of this program.

---

<sup>24</sup> The table is not reported and is available upon request.

<sup>25</sup> All numbers here are in 2011 value.

Taking into consideration that the economic return to education in rural China involves the decisions of off-farm work and migration, de Brauw and Rozelle (2006) find that the average economic return to a year of education in rural China during the 1990s was 10.5 percent for individuals younger than 35, and the average hourly wage rate for individuals younger than 35 was 2.69 yuan. Then the average monthly wage for young workers was 448.3 yuan in the 1990s.<sup>26</sup> Combined with our estimates, the monetary value of *annual* educational benefit of the rural drinking water treatment program for the youth is therefore 610 yuan, or around 87.1 dollars.<sup>27</sup> Since the average cost of the program is slightly less than 30 dollars per capita (Meng et al., 2004), the *annual* return from this investment would be around 290 percent. Given the durability of the water plant for a village, and other benefits associated with the program such as those on health (Zhang 2012), the total returns must be significantly larger. The construction of the program thus proves to be highly cost effective.

Keep in mind that the estimated return of this program does not happen instantaneously, but rather after more than 9.1 years of exposure to this program on average. In addition, we do not take into account the maintenance costs for water plants and pipelines. Such costs, however, must be smaller than the initial construction costs. Taking such additional costs into account should not alter the soundness of the water treatment program. It is also useful to know that in the rest of this paper we find the estimates of the effect of Water Plant on grades completed to be bounded between 0.44 (when village dummies are controlled for) and 1.29 (for girls using the baseline specification), which translate into annual returns to the program from 118% to 347% (again ignoring the annual maintenance costs and other benefits of the water plant).

### **A Placebo Test for Older Cohorts**

The regressions so far demonstrate positive impacts of the rural drinking water program on youth education. However, since the program was not randomly assigned and implemented by the local governments, our baseline estimates could be inconsistent due to unobserved variables

---

<sup>26</sup> The monthly payment is 448.3, calculated as  $2.69 \times 8 \text{ (hours per day)} \times 20.83 \text{ (days per month)}$ . Here 20.83 work days per month is stipulated by Ministry of Labor and Social Security of People's Republic of China (2008).

<sup>27</sup> That is,  $610 = 448.3 \times 12 \times 10.5\% \times 1.08$ . Here the exchange rate is assumed to be 7 yuan/dollar, roughly the average in our sample period.

influencing both the construction of water plants and youth education. A useful falsification test to shed light on omitted variable bias is to examine whether a plant water program is significantly related to the education of older individuals whose education is unlikely to be affected by this program. If other factors lead to the plant water effects, we would likely see the water program to be significantly related to the education of older individuals as well.

To implement this placebo test, we construct a sample of males who had passed 30 years of age when the program was implemented in their villages. We exclude females from the sample because their current residence locations likely differ from where they lived when they were in school. In contrast, males are relatively stable due to the social norms of male-biased inheritance in China. Similar to the construction of the youth sample, we only keep the observation of an individual at his/her oldest age in the CHNS data. In total the placebo sample has 2,708 individuals. The outcome variables, the same as before, reflect their final educational achievement.<sup>28</sup>

For the adults from these older cohorts, their current household information is different from that when they made decisions on schooling. We thus do not have their household characteristics during schooling ages, and have to resort to more limited control variables here. Table 3 shows the regression results. As expected, none of the “treatment effects” of plant water are statistically significant. The placebo test thus renders support for our identifying strategy.<sup>29</sup>

## **Endogeneity of Water Plant**

---

<sup>28</sup> The average grade of education completed in this sample is 7.33 years. About 50 percent of them graduated from a middle school and 15.5 percent from a high school. This placebo sample is comparable to the baseline one in their environment. Their means of the distances to schools are quite similar in magnitudes. For example, the distance to a middle school in the placebo sample is 2.35 km, as compared to 2.03 km in the baseline sample.

<sup>29</sup> The model specification for this falsification test differs slightly from the baseline specification--some individual and household characteristics (e.g. the relationship to household head, household size and income) are not controlled for due to the lack of data when those older cohorts were in school age. However, the estimates of “the treatment effects” for this placebo sample should not change much since the program implementation is relatively exogenous and, thus, uncorrelated to those characteristics at the individual and household level. To ensure that the insignificance of the “treatment effects” in this falsification test is not caused by the change in the set of control variables, we apply the same specification here to the male youth aged between 18 to 25. The regression results show that the estimated coefficients remain similar to our baseline ones. For instance, the coefficient of Water Plant on grades completed is 0.87 with the extra controls, and 0.9 without them. The results on the other two discrete outcomes are also very similar.

The country-year fixed effect regressions may be inconsistent when there is endogenous program placement. Consistent estimation of the causal treatment effects requires  $E(T_{vt}\varepsilon_{ivt}|X_{ivt}, w_{ct})$  to be zero. In other words, the installation of water plants and pipelines needs to be exogenous conditional on  $X_{ivt}$  and  $w_{ct}$ . By employing the county-year fixed effects, we are able to capture the unobservable at the county-year levels. However, if some unobservables that vary within counties across years affect both the timing and the location of the program and simultaneously affect education at the village level, we still face the thorny issue of endogeneity for  $T_{vt}$ , which results in inconsistent estimates of the treatment effect.

A simple way to shed light on the endogeneity of the treatment placement is to examine whether pre-treatment characteristics between the treatment and the comparison group are similar. We thus construct the treated group as those villages that experienced a change in the treatment status,<sup>30</sup> and the comparison group as the never-treated group. Since we do control for country-year fixed effects in our base regressions, all pre-treatment characteristics in this exercise are net of the influence of these dummy variables. The results, in Table A3 of the appendix, show that the pre-treatment characteristics are very similar between the treatment and the comparison groups. Thus, the never-treated and the treatment-status-changing villages are similar in pre-treatment characteristics, rendering support to our county-year fixed effects specification. A caveat here is that this test cannot shed light on whether the always-treated villages have similar pre-treatment characteristics due to the lack of data in pre-treatment periods.

Even though we find some support for the country-year fixed effects specification, we cannot completely rule out the potential endogeneity of the treatment. As a way to deal with potential endogeneity, we instrument  $T_{vt}$  with the topographic characteristics of villages (i.e., whether the village is geographically flat, hilly or mountainous), which influence the costs of the construction of water plants and pipeline systems in several ways. Fixed costs are higher in non-flat areas since it is more difficult to lay pipes, and high-pressure water pumps must be installed to deliver water. Similarly, variable costs are also higher in non-flat areas as a large amount of electricity is needed to pump water from plants to villages.

---

<sup>30</sup> Since always-treated villages do not have pre-treatment characteristics, we cannot include them in the treatment group for the pre-treatment test.

Our key identifying assumption is that, conditional on demographic characteristics, household income, accessibility of schools and the county-year fixed effects, the topographic characteristics of the villages should affect people's education only through the water treatment program. Topography, or land gradient, has been shown in the literature as affecting agricultural productivities (Udry, 1996), crop types (Qian, 2008) and infrastructure construction (Duflo and Pande, 2007; Donaldson, 2010; Dinkelman, 2011). It is plausible that these factors may affect individuals' educational status—other than through the water treatment program—mainly through the household income. Therefore, controlling for household income in the regressions can help satisfy the exclusion restriction when using the villages' topography as the instrument. Based on the description of the topography of a village in the CHNS survey, we construct a dummy variable of a village being non-flat as the instrument for Water Plant.<sup>31</sup> The F-statistics for the excluded instrument in the first stage is 21.23, which suggests that our topographic instrumental variable is not weak.<sup>32</sup> Note that the village's topography was only recorded in the survey in 1991. As a result, the sample size for IV regressions is smaller since the topography of the newly-added villages after 1991 is not available.

Table 4 presents the IV estimates along with our baseline estimates. The qualitative conclusions from the IV estimates are similar to the baseline estimates, although the magnitudes of IV estimates double those of the OLS ones. For instance, Water Plant has a coefficient of 2.41 (statistically significant at the one percent level) while the baseline estimate is 1.02 (and statistically significant at the one percent level). The IV coefficient for graduation from a middle school remains positive and significant, with a larger magnitude (0.359 in IV vs 0.128 in FE). The coefficient for graduation from a high school remains positive and also statistically significant.

We cannot reject the null hypothesis that the IV and the FE coefficients are equivalent jointly. In particular, we conduct a bootstrapped Hausman test for each outcome variable with the

---

<sup>31</sup> The survey allows us to construct slightly finer instrument (i.e., distinguishing hilly or mountainous areas). But it is often not obvious to distinguish between being hilly or mountainous, and the results using two geographical instruments actually lead to weaker instruments as shown in the F-statistic. The IV estimates of the treatment effects using the two IVs are still positive and large, but less precisely estimated.

<sup>32</sup> As recommended by Baum, Schaffer and Stillman (2007), we compare the statistic to the rule of thumb (10) as well as the Stock-Yogo critical values because critical values for the heteroskedastic-robust Kleibergen-Paap Wald rk F Statistic of the test have not yet been calculated. The Stock-Yogo weak ID test critical value is 16.38 for 10% maximal IV size.

null hypothesis that the OLS and the IV estimates are statistically equal (Cameron and Trivedi, 2005).<sup>33</sup> The p-values generated from these tests are all above 0.5, indicating a failure to reject the null hypothesis. The IV results thus render support to our county-year FE results of significant and positive effects of access to plant water on education.<sup>34</sup> In light of the Hausman test results we shall stick to the county-year FE specification in all future specifications since it is more efficient.

### **Controlling for More Village Characteristics**

When choosing the baseline specification, we opt to control for the county-year FEs rather than the village FEs for the following key reason. With the slow rollout of the water program, only 45 out of 174 villages in the sample experienced changes in their treatment status during the sample period. When we use village fixed effects, our identification comes from the before-after changes in the outcomes for these 45 villages (after controlling for other covariates) when Water Plant changes from zero to one. This strategy thus risks throwing the baby out with the bath water: the variations in education and water quality for the always-treated and the never-treated villages (within a county-year cell), though quite informative, are totally ignored in identifying the treatment effects.

However, it is still a legitimate concern that omitted characteristics at the village level can cause a spurious correlation between Water Plant and education. To check this possibility we proceed in three steps. *First*, we control for more village characteristics, including the average income and the Gini coefficient of the village. This partially deals with the issue of omitting key village characteristics while preserving the full sample. The results are in Table A2. The estimated treatment effects are qualitatively similar to the baseline ones. The magnitudes drop slightly. The effect of Water Plant changes from 1.08 in the baseline specification to 0.96 here.

---

<sup>33</sup> The bootstrapped Hausman tests are conducted as follows: (i) estimate OLS and an IV estimates from a bootstrap subsample with the village as the resampling cluster; (ii) repeat this process 1000 times to calculate the standard errors of those estimates; (iii) conduct the Hausman test by using the estimated coefficients using the whole sample and the standard errors obtained in step (ii).

<sup>34</sup> We have also tried using the non-flat dummy interacted with all wave dummies as instruments. The qualitative results are similar to the results with the simple non-flat dummy IV, with significant treatment effects on grades completed (2.22 years) and the dummy variable of graduating from middle school (34 percentage points). However, the first-stage F statistics become smaller (10.11) which is over 10 but less than the Stock-Yogo weak ID test critical value (20.53) for 10% maximal IV size.

*Second*, we deal with the concern that we may have omitted local village-level labor market conditions in our baseline regressions. The youth's education decisions are usually made jointly with their labor supply decision, so their pursuit of education may be influenced by local labor market conditions, such as wage rates or job vacancies. In the baseline regressions, we control for county-year dummies which capture the overall labor market conditions at the county level, but we ignore labor market variables at the village level. Whether this omission can bias the estimates of the treatment effects depends on whether these variables are correlated with Water Plant. In its community-level survey, CHNS contains some basic facts of the local labor market, including daily wages for male, female and construction workers, respectively, and the major occupations that the local residents are engaged in. In order to measure village-level real wages, we normalize all nominal wages by the prices of commodities mostly consumed daily.<sup>35</sup>

Table 5 presents regression results of village wage rates (normalized by the local rice price index) on Water Plant.<sup>36</sup> The local real wage rates do not have statistically significant correlations with the water program. In separate regressions, we also control for the travel costs (proxied by the road conditions around villages), which may affect the relative attractiveness of local jobs and thus affect the work-school choice. Even after we control for the proxies of travel costs, there remains no correlations between Water Plant and local real wages. Omitted time-varying village-level labor market conditions thus cannot affect our estimate of the treatment effect.

*Third*, we control for village fixed effects. The inclusion of village fixed effects has two adverse consequences: (i) it results in a dramatic drop in the useful sample for identification, and (ii) the 45 treatment-status-changing villages experienced a shorter history with treated water, which makes it harder for a new water plant to leave a mark on education. This is likely problematic in light of our findings presented later: the most pronounced effect of water treatment is observed for those youth who had their first exposure to water treatment between age zero and two. Relying exclusively on the treatment-status-changing villages would force us

---

<sup>35</sup> Since we do not have the village-level price index, using the price of commodities mostly consumed daily in the village is the best alternative. Since we have already controlled for county-year dummies, all county level price variations have been controlled for.

<sup>36</sup> In addition to local rice price, we have also tried employing the prices of flour, oil, pork, peanut oil, and eggs to adjust the normal wages. All such adjusted wages are not statistically significantly correlated to the water improvement program.

to rely more on youth who got exposed to plant water at age older than 2. As a consequence, the final sample would have a smaller share of individuals who got exposed to treated water at age zero to two, and the average treatment effects would be smaller simply because it omits (to a greater extent) those sample who benefits the most from treated water.<sup>37</sup> As a result of this consideration, we expect to find significantly weaker treatment effects when controlling for village fixed effects as compared to using county-year fixed effects. This weakening of treatment effects must be stronger for outcomes that take a longer term to materialize--the dummy of graduating from a middle or high school.

Presenting the regression results with both county-year and village FE controls, Table 6 confirms our conjectures. Controlling for village FEs decreases the magnitudes of the treatment effects for all of the outcome variables. The gain in grades of education completed drops from 1.08 to 0.44 years, and the estimate of the likelihood of being a graduate of a middle school and a high school change from 13 to 5 percentage points and from 14.7 to 3.2 percentage points, respectively. Controlling for village FEs also makes the estimates of the effects on two dummy outcome variables statistically insignificant, but the coefficient of Water Plant for grades of education completed stays statistically significant the five percent level. Thus, even with the variation coming from 26% of the original sample villages and only focusing on their before-after comparison, and ignoring more those individuals who began their treatment at early childhood, we still find a significant effect of Water Plant on grades attained. The coefficient of 0.437, representing the lowest bound for the water plant effect in all our specifications, still translates into a huge annual return to investing in a water plant: 118%.<sup>38</sup> This again supports our conclusion that the water plant for rural people in China has been highly cost-effective.

### **Does the Plant Water Effect Reflect Health Improvement or Time Saving?**

---

<sup>37</sup> In particular, for data of wave 1997 and after, those individuals who got exposed to treated water at age zero to two would be excluded from the treated sample.

<sup>38</sup> We obtain this figure as follows: The monthly wage is constructed as 2.69 yuan (i.e., hourly wage rate) \* 8 hours \* 20.83 days per month, or 448.3 yuan. The annual returns from having a water plant is 448.3 (i.e., month wage) \* 12 months \* 10.5% (i.e., educational returns for an additional year of education) \* 0.437 (i.e., the education effect of having a water plant). Changing this number into dollars (by dividing it by 7) and dividing the dollar amount by the cost of the plant construction (30 dollars), we obtain 118%.



So far we have found significant effects of access to plant water, and we interpreted it as the effects of improved water quality. However, the water-education link resulted from the program could reflect water *access* rather than water *quality*. Kosec (2014), for instance, finds strong benefits of better access to piped water in African countries. Better water access could save a household's time spent on fetching water, and allows family members to enjoy more *quantity* of water. If fetching water is mainly shouldered by the youth, access to plant water through pipelines allows the youth to have a larger time budget to attend school and to complete school tasks.

We thus examine whether the link between plant water and education is due to the omission of water access. In CHNS, households' water access is categorized into: (i) in-house tap water, (ii) in-yard tap water, (iii) in-yard well, and (iv) other places. The first three options should be considered as optimal water access (Esrey 1996, Mangyo 2008).<sup>39</sup> We thus create a dummy variable of optimal water access which is one when water is accessible on the household premise.

We rerun our baseline regressions by controlling for the dummy of optimal water access for a household. Alternatively we limit the sample to the individuals who have optimal water access for all the sample years. If the main channel of the Water Plant effect is through water access, the coefficient of Water Plant should dwindle to zero, and the optimal access dummy to be significant and positive with a large magnitude. The regression results are in Table 7.

The coefficients of Water Plant are only slightly smaller than in the baseline (when we do not control for water access) and remain highly significant. The treatment effect on grades completed is 1.01 year, as compared to 1.08 year in the baseline regression. When restricting our sample to the individuals always having access to water in their premises, the treatment effect of 1.04 years is almost identical to the baseline results. Thus, the effect of plant water program on education is not due to improved water access and its corresponding time-saving effect. This, couples with the findings in Zhang (2012) that plant water improves health status for both adults and children, suggests that improved health due to plant water (and their indirect effects on

---

<sup>39</sup> Manyo (2008) also uses this definition when studying the health effects of water access in China.

household income) is more likely the explanation for the educational effects of access to plant water.<sup>40</sup>

### **Household Income and Treatment Effect**

Does the treatment effect depend on household income? The answer to this question might shed light on the distributional impact of the water treatment program.<sup>41</sup> To answer the question, we divide the sample into two groups according to their baseline household incomes, each accounting for half of the income distribution. Table 8 shows the regression results for each outcome and for each income group. Overall, the coefficients for the high income group are reasonably close to those for the low income group in magnitudes and all of the coefficients are statistically significant at the one percent level. For example, the effect of plant water on grades of education completed is 0.984 for the poor vs. 1.043 for the rich. Although here the impacts are slightly stronger for the rich than the poor, this pattern diminishes—we find similar effects across groups--when we run regressions for three income groups. Our results thus suggest that water treatment in rural areas benefited the poor and the rich similarly.

### **Time of Exposure and Treatment Effects**

---

<sup>40</sup> Note that the coefficients of the optimal water access are significant in regressions with outcomes of grades of education completed and the probability of graduating from a middle school, while not with the outcome of graduating from a high school. This finding fits the institution of China's educational system. Rural students usually commute from home to schools when they are in primary or middle schools since the distances to those schools are not far (The average distance to a primary school and a middle school are 0.57 kms and 2.03 kms, respectively). In this case children can benefit from the availability of water in their premises since their time to fetch water can be saved. In contrast, a high school is normally located far away and our data shows that the average distance to a high school is 8.5 kms. Rural students in high schools are not able to travel between home and schools every day. They either live with relatives or rent a place close to schools. Quite often high schools also provide accommodation for their students as well. Those students do not do household chores such as fetching water. This observation explains why better access to water does not have a statistically significant impact on the children's probability of graduating from a high school.

Our finding that optimal water access improves education is also consistent with Maimaitwe and Siebert (2009), who use CHNS data to find that poor *access* to water hurt girls' education, but not boys (age between 6 and 19), which also imply a positive average effect of water *access* on education. Their definition of optimal water access however differs from ours; they exclude but we and Manygo (2008) include "in-yard well" in the optimal water access category. Our results are thus not strictly comparable.

<sup>41</sup> Galiani et al. (2005) and Kosec (2014), for instance, both find that private sector participation in water supply tends to benefit the poor more than the rest.

So far we have not distinguished the treatment effect among individuals who got treated at different ages. However, interventions at different stage of life could result in dramatically diverse returns (Heckman, 2008; Cunha et al. 2006, 2010; Almond and Currie, 2011; Campbell et al., 2014). The nutrition literature, for instance, emphasizes that early childhood period is critical, and that poor nutrition in this period is hard to be compensated by better nutrition in later years (Martorell 1995, 1997; Martorell et al. 1994; Maccini and Yang, 2009). Part of the reasons why the return to human capital investment during this early childhood period (i.e., from birth to age 5 when the brain develops rapidly) is the highest is that early childhood intervention fosters cognitive and non-cognitive skills (Heckman, 2008). While there are studies on the long-term impacts of early childhood health or nutritional status (Glewwe and King, 2001; Glewwe et al., 2001; Case et al. 2002; Currie and Stabile 2003; Currie et al. 2010), or home environment, or toxic exposure (Nilsson 2009), or exposure to health shocks such as infections or drought (Hoddinott and Kinsey, 2001; Bleakley 2007, Chay et al. 2009, Case and Paxson 2009),<sup>42</sup> studies on the long-term effects of early childhood exposure to treated water on education are rare. It is therefore important to examine whether the earliest age that an individual is exposed to plant water matters differently from later ages.

To proceed, we calculate the earliest year of exposure to plant water for each individual in treated villages. Due to the lack of information, we delete from our sample those individuals in the always-treated villages who were born before the first wave when the villages were covered in CHNS. As a result, 665 individuals drop out of our sample, and the sample is now 4,064 observations. Then we explore the effects of exposure time non-parametrically for the outcome of grades of education completed (see Table 9). In Column (1), we just rerun the baseline regression by using this restricted sample. Not surprisingly, the treatment effect decreases to 0.65 years since we exclude the observations that have used plant water for the longest time in our sample and whose educational attainment may benefit the most from this water program. Nevertheless, the treatment effect stays statistically significant at the one percent level.

In Column (2), we create a series of dummy treatment variables to indicate individuals' earliest exposure age to plant water and rerun the regression with those treatment variables. The

---

<sup>42</sup> See Almond and Currie (2011) for a summary of the literature on returns to early childhood human capital investment or environment.

result clearly shows that the treated plant water has the greatest impact when it is available in early childhood, in particular, at age two and younger. For example, on average a child's final educational achievement increases by 1.69 years if he/she starts to use plant water before 2 years old. And this impact decreases to 0.4 to 0.6 year when the earliest exposure age is older than age 2. The treatment effect stays more or less flat after age two. In Figure 4 we plot the coefficients and their 90 percent confidence intervals of those dummy treatment variables indicating a child's earliest exposure age.<sup>43</sup> The finding of the highest return at the pre-school stage (i.e., birth to age 2) is consistent with the literature that suggest that the largest return to human capital investment comes at the early childhood (Heckman, 2008; Almond and Currie, 2011). We add to this literature by showing that exposure to treated plant water also exhibit highest return to education at the early childhood stage.<sup>44</sup>

### **Heterogeneous Treatment Effects across Gender**

During the sample period, rural Chinese girls on average are less educated than boys. In our data, the average grades of education completed is 8.54 years for girls versus 8.85 years for boys, and 63.4 and 15.9 percent of girls are graduates of middle and high schools, respectively, as compared to 68.1 and 16.8 percent for boys. Can health improvement programs such as the rural drinking water treatment program significantly reduce the gender gap? The brawn theory of division of labor (Pitt et al., 2012) would imply so. To answer this question, we explore the heterogeneous treatment effects for boys and girls.

Table 10 presents the results for boys and girls separately. In general, the treatment effects on girls are substantially larger than those on boys. For example, the grades completed

---

<sup>43</sup> We have also explored the treatment effects of this water program in a parametric method. First, we construct a continuous variable which equals to  $(26 - \text{earliest exposure age})/10$  and coded this variable as zero for all of the individuals who do not have access to plant water before 25. This definition of this variable implies that exposure to plant water after 26 is zero. We then substitute the polynomial terms of exposure age up to the 3<sup>rd</sup> order for the simple dummy treatment variable—Water Plant—in the baseline regression with the outcome of grade of education completed. The regression results confirm that the largest impact is felt at the early childhood years, after which the impact is much more contained.

<sup>44</sup> A counter-intuitive finding here is that the treatment effects of treated water on individuals who were first exposed to treated water between age 20 and 25 remain positive and significant. Our explanation is that late enrollments often happen in undeveloped areas. For example, children at 12 years old are supposed to have finished primary schools or at least are at the 6<sup>th</sup> grade in school. However, the CHNS data shows that more than 35 percent of the twelve-year olds study at the 5<sup>th</sup> grade, around 13 percent at the 4<sup>th</sup> grade, and 6 percent at the 3<sup>rd</sup> grade or below.

increase by 0.87 year for boys but 1.29 years for girls when their villages gain access to treated plant water. And the treatment effect on being graduates from a middle school for girls is 15.1 percentage points, around 37 percent greater than that for boys (11 percentage points). Note that the girl advantage in the water plant effect, 0.42 years of schooling, represents more than 100% of the girl disadvantage in schooling attainment in our sample (i.e., 0.31). The rural drinking water program thus contributed significantly to gender equality in education in the treatment regions in the countryside.

The stronger effect of our water program on girls than on boys resonates well with some recent empirical and theoretical papers that find stronger effects of health programs on girls.<sup>45</sup> Miguel and Kremer (2004) find that the spread of deworming drugs among schools in Kenya increased the school attendance instantly for both boys and girls, but this effect lasted in the second year only for girls. Maluccio et al. (2009) shows a significant increase in schooling attainment only for girls but not for boys in Guatemala after they were both treated by nutritional supplements for three years. Maccini and Yang (2009) find that Indonesian women enjoyed long-term benefits in terms of education and other socioeconomic outcomes when they were exposed to favorable weather shocks when they were infants, but this is not true for Indonesian men.

The theoretical rationale for stronger responses by girls from health improvement is provided by Pitt et al. (2012), which introduces gender differences in the level of brawn and responsiveness of brawn to nutrition. As found in the biomedical literature, biologically boys have more brawn than girls, and boys' brawn grows more when their nutritional status has been improved. Thus young male have comparative advantage and are more likely to work in brawn-intensive (as opposed to skill-intensive) jobs. A health intervention reduces morbidity and improves an individual's nutritional status, through which boys gain more brawn than girls. Such a health shock increases young men's comparative advantage in working in brawn-intensive occupations, and, thus, raises the opportunity costs of their schooling. This naturally tilts young men toward work and young women toward schooling (i.e., skill acquisition)—at least relatively—and this stronger pull for physical work is amplified by what has been going on in

---

<sup>45</sup> See Pitt et al. (2012) for a summary of other studies finding differential investment in and returns to human capital investment.

China's labor markets: over the past 3 decades, a huge amount of migrants have moved to cities, with rural males tend to work in brawn-intensive jobs such as construction, laborers, and so on. The framework of relative comparative advantage in brawn-intensive jobs by males thus explains why young men may choose to work more and enjoy lesser gains in education in response to health interventions.

To test the hypothesis of gender differences in the level of brawn and responsiveness of brawn to nutrition, we replicate Zhang (2012) and run regressions that relate boys' and girls' health status to Water Plant and other control variables. Here we are trying to understand the health effects of treated plant water for those young males and females, so we track their nutritional status recorded in CHNS before 25 and explore whether access to plant water improves their health and whether this improvement differs by gender. The sample size used for this analysis is 11,169 observations.<sup>46</sup> The results are in Table 10. Both boys and girls gain improvements in their height and body mass (i.e., weight/height) after they have been exposed to plant water. The key finding is that boys' body mass has increased by 0.55 kg/m, which is much larger than such effect on girls (0.39 kg/m). Moreover, there is evidence that the same increase in body mass translate into more gain in strength for male than for female (Pitt et al. 2011). Thus, boys gain more brawn than girls after treated plant water becomes accessible. Our findings of gender-specific brawn-responsiveness to plant water thus also offer support to the brawn story in Pitt et al. (2012).

The brawn theory of gender division of labor has interesting implications about how an elder sibling's gender identity affects education of younger siblings within a family through the income channel. When treated plant water improves health of young people in a household, this theory implies that the boys are relatively more likely to choose to work and girls more likely to continue schooling. As a result, male youth are more able to contribute to their household financial resources and provide better support to the schooling pursuit of their younger siblings. We thus expect that an individual with an elder brother gains more in education from

---

<sup>46</sup> To construct the sample to examine the health effects of plant water between boys and girls, we track the measures of nutrition of our baseline individuals from age 0 to 25 recorded in CHNS data. Thus, the final sample is panel data of the individuals showing up in the sample for analyzing the effects on education, which is larger than the cross-sectional data we use in the baseline analysis.

improvement of drinking water quality because of the larger increase in household financial resources earned by the elder brother than those with an elder sister.

Table 11 tests this conjecture by presenting the treatment effects by the gender of the elder sibling. For our regressions, our sample consists of the individuals who are children of household heads and also have an elder brother(s) or sister(s). For the gender of the elder sibling, we divide the sample based on the gender of the eldest sibling. The treatment effects in terms of grades completed for individuals with an elder brother more than double those for individuals with an elder sister (i.e., 1.39 vs. 0.57 years), rendering further support to the brawn theory of the gender division of labor.

## **5 Conclusion**

In this paper we study the impacts on education of a major rural drinking water treatment program in China, which aims to build up water purification plants and pipelines to provide treated safe drinking water for rural residents. We find that post-high-school-age young people in villages with access to treated plant water have better education than those without such access: the youths' grades of education completed improve by 1.08 years and their likelihoods of graduating from middle and high schools both increase by 13 and 14.7 percentage points, respectively. The results are obtained after controlling for county-year dummies (and therefore local educational policies and resources), household characteristics, and village characteristics including the distances to schools. The qualitative results remain robust after dealing with the endogeneity of the water treatment program, considering local labor market conditions, controlling for water access (i.e., whether a household has access to water in its premise), and estimating the effects by income groups and gender. The main channel through which plant water coverage benefits youth education is improved health of the youth themselves and partially increased household income resulted from improved health of other household members and early entry into brawn-typed jobs of their elder brothers if any, but not time saving due to better water access. Our placebo test on adults' education also supports our identifying assumption. Three findings render support to the brawn-based theory of gender division of labor in Pitt et al. (2012): the female benefits much more from water treatment than the male in terms of schooling

attainment (1.29 vs 0.87 years); youth with an older brother benefit more in education attainment than youth with an older sister; and boys gain more in body mass than girls after water treatment. The brawn story proves to be quantitatively important: the water treatment program completely wipes out the gender gap in education in the villages with access to treated water. Interestingly, rural youth who began their exposure to treated water in their early childhood (i.e., 0-2 years of age) benefited the most from the water treatment program, consistent with the recent literature that emphasizes the critical importance of early childhood in terms of investment in human capital and health (Cunha et al. 2006; Heckman 2008; Almond and Currie, 2011). Our back-of-envelope computation of the cost-benefit of the rural water treatment program suggests that the program is highly cost-effective even under our most conservative estimate.

Our results suggest that some basic infrastructure programs such as the provision of safe drinking water can significantly increase rural educational level, therefore potentially contributing to reducing income inequality between rural and urban residents. Our findings also suggest that water treatment programs have the potential to dramatically reduce gender education gap in rural China. Our results echo recent findings that highlight the critical importance of investing in young people at the early childhood stage.

It seems that careful analyses of how Chinese fiscal resources are spent can be quite useful. The fiscal revenues and expenditures in China are growing faster than its GDP growth. Yet we know very little about the cost-effectiveness of various types of government spending in China. For this rural drinking water program, the return has been huge. What about the cost effectiveness of other programs such as those on highways or high-speed trains on which the government has spent much more than on water treatment? More careful empirical work is clearly needed to guide the allocation of fiscal resources in China and elsewhere.



## References

- Alderman, Harold; Jere R. Behrman; Victor Lavy and Rekha Menon. 2001. "Child Health and School Enrollment: A Longitudinal Analysis." *The Journal of Human Resources*, 36(1), 185-205.
- Alderman, Harold; John Hoddinott and Bill Kinsey. 2006. "Long Term Consequences of Early Childhood Malnutrition." *Oxford Economic Papers*, 58(3), 450-74.
- Almond, Douglas. 2006. "Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population." *Journal of Political Economy* 114, 672-712.
- Almond, Douglas, and Janet Currie 2011. "Human Capital Development before Age Five." In *handbook of Labor Economics*. Vol. 4B, ed. Orley Ashenfelter and David Card, Chapter 15, 1315-1486. North Holland: Elsevier.
- Audibert, Martine, 1986. "Agricultural Non-wage Production and Health Status: A Case Study in a Tropical Environment." *Journal of Development Economics*, 24(2), 275-91.
- Baum, Christopher F; Mark E., Schaffer and Steven Stillman. 2007. Enhanced Routines for Instrumental Variables/GMM Estimation and Testing. *Stata Journal*, 7(4), 465-506.
- Bleakley, Hoyt. 2007. "Disease and Development: Evidence from Hookworm Eradication in the American South." *The Quarterly Journal of Economics*, 122(1), 73-117.
- Bleakley, Hoyt. 2010. "Health, Human Capital, and Development." *Annual Review of Economics*, 2, 283-310.
- Bobonis, Gustavo J.; Edward I Miguel and Charu Puri-Sharma. 2006. "Anaemia and School Participation," *eSocialSciences*.
- Braudel, Fernand. 1982. *Civilization and Capitalism, 15<sup>th</sup>-18<sup>th</sup> Century*. New York: Harper & Row.
- Cai, Hongbin, Hanming Fang, Lixin Colin Xu. 2011. "Eat, Drink, Firms, Government: An Investigation of Corruption from Entertainment and Travel Costs of Chinese Firms." *Journal of Law and Economics* 54, 55-78.
- Cameron, Colin and Pravin Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge University Press.
- Case, Anne, Darren Lubotsky, Christina Paxson. 2002. "Economic Status and Health in Childhood: The Origins of the Gradient." *American Economic Review* 92(5), 1308-1334.
- Case, Anne, Christina Paxson. 2009. "Early Life Health and Cognitive Function at Old Age." *American Economic Review Papers and Proceedings* 99(2), 104-109.
- Center for Health Statistics and Information, Ministry of Health of the People's Republic of China, 2009. *Research on Health Services of Primary Health Care Facilities in China*. Peking Union Medical College Press, Beijing.
- Clarke, George R.G., Katrina Kosec, Scott Wallsten. 2009. "Has Private Participation in Water and Sewerage Improved Coverage?" *Journal of International Development* 21(93), 327-361.
- Cunha, Flavio, James J. Heckman, and Susanne M. Schennach. 2010. "Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Econometrica* 78(3), 883-931.

- Cunha, Flavio, James J. Heckman, Lance Lochner, Dimitriy V. Masterov. 2006. "Interpreting the Evidence on Life Cycle Skill Formation." In *Handbook of the Economics of Education*, Volumen 1, eds. By Eric A. Hanushek and Finis Welch. Elsevier.
- Currie, Janet, Mark Stabile. 2003. "Socioeconomic Status and Child Health: Why is the Relationship Stronger for Older Children?" *American Economic Review* 93(5), 1813-1823.
- Currie, Janet, Mark Stabile, Phongsack Manivong, Leslie L. Roos. 2010. "Child Health and Young Adult Outcomes." *The Journal of Human Resources* 45(3), 517-548.
- Chay, Kenneth, Jonathan Guryan, Bhashkar Mazumder. 2009. "Birth Cohort and the Black-White Achievement Gap: The Roles of Access and Health Soon after Birth." NBER working paper 15078.
- De Brauw, Alan and Scott Rozelle. 2008. "Reconciling the Returns to Education in Off-Farm Wage Employment in Rural China." *Review of Development Economics*, 12(1), 57-71.
- Dickson, Rumona; Shally Awasthi; Paula Williamson; Colin Demellweek and Paul Garner. 2000. "Effects of Treatment for Intestinal Helminth Infection on Growth and Cognitive Performance in Children: Systematic Review of Randomised Trials." *BMJ*, 320.
- Dinkelman, Taryn. 2011. "The Effects of Rural Electrification on Employment: New Evidence from South Africa." *American Economic Review*, 101(7), 3078-108.
- Donaldson, Dave. 2010. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." National Bureau of Economic Research Working Paper Series, No. 16487.
- Duflo, Esther and Rohini Pande. 2007. "Dams." *The Quarterly Journal of Economics*, 122(2), 601-46.
- Esrey, Steven A. 1996. "Water, Waste, and Well-Being: A Multicountry Study." *American Journal of Epidemiology*, 143(6), 608-23.
- Esrey, Steven A, J.B. Potash, L. Roberts, and C. Shiff. 1991. "Effects of Improved Water Supply and Sanitation on Ascariasis, Diarrhoea, Dracunculiasis, Hookworm Infection, Schistosomiasis, and Trachoma." *Bulletin of the World Health Organization* 69(5), 609-21.
- Fewtrell, Lorna, Rachel B. Kaufmann, David Kay, Wayne Enanoria, Laurence Haller, and John M. Colford Jr. (2005). "Time to Focus Child Survival Programmes on the Newborn: Assessment of Levels and Causes of Infant Mortality in Rural Pakistan." *Bulletin of the World Health Organization*, 80, 271-276.
- Galiani, S., P. Gertler, E. Schargrodsky. 2005. "Water for Life: The Impact of the Privatization of Water Services on Child Mortality." *Journal of Political Economy* 113(1), 83-120.
- Gamper-Rabindran, Shkkee Khan Shanti, and Christopher Timmins. 2010. "The Impact of Piped Water Provision in Infant Mortality in Brazil: A Quantile Panel Data Approach." *Journal of Development Economics* 92, 188-200.
- Glewwe, Paul; Elizabeth M. King. 2001. "The Impact of Early Childhood Nutrition and Academic Achievement: Does the Timing of Malnutrition Matter?" *World Bank Economic Review* 15(1), 81-114.
- Glewwe, Paul; Hanan G. Jacoby and Elizabeth M. King. 2001. "Early Childhood Nutrition and Academic Achievement: A Longitudinal Analysis." *Journal of Public Economics*, 81(3), 345-68.

- Glewwe, Paul and Harry Anthony Patrinos. 1999. "The Role of the Private Sector in Education in Vietnam: Evidence from the Vietnam Living Standards Survey." *World Development*, 27(5), 887-902.
- Heckman, James J. 2008. "Schools, Skills, and Synapses." *Economic Inquiry* 46(3), 289-324.
- Hoddinott, J., and B. Kinsey. 2001. "Child Growth in the Time of Drought." *Oxford Bulletin of Economics and Statistics* 63(4), 409-436.
- Jalan, Jyotsna, Martin Ravallion. 2003. "Does Piped Water Reduce Diarrhea for Children in Rural India?" *Journal of Econometrics* 112(1), 153-73.
- Kremer, Michael, Jessica Leino, Edward Miguel, and Alix Peterson Zwane. 2011. "Spring Cleaning: Rural Water Impacts, Valuation, and Property Rights Institutions." *Quarterly Journal of Economics* 126(1), 145-205.
- Liu, Mingxing; Rachel Murphy; Ran Tao and Xuehui An. 2009. "Education Management and Performance after Rural Education Finance Reform: Evidence from Western China." *International Journal of Educational Development*, 29(5), 463-73.
- Maccini, Sharon, and Dean Yang. 2009. "Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall." *American Economic Review*, 99(3): 1006-26.
- Maimaiti, Yasheng and Siebert, Stanley, 2009. The Gender Education Gap in China: The Power of Water. IZA Discussion Paper No. 4108.
- Maluccio, John; John Hoddinott; Jere R. Behrman; Reynoldo Martorell; Agnes R. Quisumbing and Aryeh D. Stein. 2009. "The Impact of Improving Nutrition During Early Childhood on Education among Guatemalan Adults." *The Economic Journal*, 119 (537): 734-763.
- Mangyo, Eiji. 2008. "The Effect of Water Accessibility on Child Health in China." *Journal of Health Economics*, 27(5), 1343-56.
- Martorell, R. 1995. "Results and Implications of the INCAP Follow-Up Study." *Journal of Nutrition* 125, 1127S-1138S.
- Martorell, R. 1997. "Undernutrition During Pregnancy and Early Childhood and its Consequences for Cognitive and Behavioral Development." In M.E. Young (ed.), *Early Child Development: Investing in Our Children's Future*. Amsterdam: Elsevier, 39-83.
- Martorell, R., K.L. Khan, D.G. Schroeder. 1994. "Reverse Ability of Stunting: Epidemiological Findings in Children from Developing Countries." *European Journal of Clinical Nutrition* 48, 45S-57S.
- Martorell, R; J P Habicht and J A Rivera. 1995. "History and Design of the Incap Longitudinal Study (1969-77) and Its Follow-up (1988-89)." *Journal of Nutrition*, 125, 1027S-1041S
- Meng, S.; J. Liu and Y. Tao. 2004. "Water Supply and Sanitation Environment in Rural China: Promote Service to the Poor." Paper presented at the Poverty Reduction Conference, Shanghai
- Miguel, Edward and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica*, 72(1), 159-217. Ministry of Health of the People's Republic of China, 2007. *Chinese National Health Statistics 2007*. Peking Union Medical College Press, Beijing.
- Ministry of Health of the People's Republic of China, 2007. *Sanitary Standard for Drinking Water Quality GB5749-2006*. Standard Press of China, Beijing.

- Ministry of Labor and Social Security of People's Republic of China, 2008. *Circular of the Ministry of Labor and Social Security on the Issues Concerning the Average Monthly Working-time in a Year and Converting the Payment of Employee*, [http://www.molss.gov.cn/gb/zxwj/2008-01/10/content\\_219002.htm](http://www.molss.gov.cn/gb/zxwj/2008-01/10/content_219002.htm) (accessed on Jan 1, 2013).
- National Bureau of Statistics of China, 2007. *China Statistical Yearbook 2007*. China Statistics Press, Beijing.
- Nguyen, Quynh. 2010. "Essays in Empirical Microeconomics on Economic Development." Doctoral Dissertation, Department of Economics, University of Maryland (College Park).
- Pitt, Mark M.; Mark R. Rosenzweig; and Mohammad Nazmul Hassan. 2012. "Human Capital Investment and the Gender Division of Labor in a Brawn-Based Economy." *American Economic Review*, 102(7), 3531-60
- Pollitt, Ernesto; Kathleen S. Gorman; Patrice L. Engle; Reynaldo Martorell; Juan Rivera; Theodore D. Wachs and Nevin S. Scrimshaw. 1993. "Early Supplementary Feeding and Cognition: Effects over Two Decades." *Monographs of the Society for Research in Child Development*, 58(7), i-118.
- Qian, Nancy. 2008. "Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance." *The Quarterly Journal of Economics*, 123(3), 1251-85.
- Nguyen, Quynh. (2010). *Essays in Empirical Microeconomics on Economic Development*. Unpublished doctoral dissertation, University of Maryland, College Park, US.
- Singh, Inderjit; Lyn Squire, and John Strauss. 1986. "The Basic Model: Theory, Empirical Results, and Policy Conclusions." in *Agricultural Household Models*. Baltimore: Johns Hopkins University Press.
- Strauss, John. 1986. "Does Better Nutrition Raise Farm Productivity?" *Journal of Political Economy*, 94(2), 297-320.
- Tan, Song-hua. 2003. "Status quo, Difficulties and Countermeasures of the Development of Rural Education in China." *Peking University Education Review* (in Chinese), 1(1), 99-103.
- Thirumurthy, Harsha; Joshua Graff Zivin and Markus Goldstein. 2008. "The Economic Impact of Aids Treatment." *Journal of Human Resources*, 43(3), 511-52.
- Thomas, D., Frankenberg, E., Friedman, J., Habicht, J-P., Hakimi, M., Ingwersen, N., Jaswadi, Jones, N., McKelvey, C., Pelto, G., Seeman, T., Sikoki, B., Smith, J.P., Sumantri, C., Suriastini, W., and Wilopo, S. (2006). "Causal effect of health on labor market outcomes: Experimental evidence." CCPR working paper #2006-70. Los Angeles: California Center for Population Research, University of California.
- Thomas, Duncan and John Strauss. 1997. "Health and Wages: Evidence on Men and Women in Urban Brazil." *Journal of Econometrics*, 77(1), 159-85.
- Tonglet, Rene, Katulanya Isu, Munkatu Mpese, Michele Dramaiz, and Philippe Hennart. 1992. "Can Improvements in Water Supply Reduce Childhood Diarrhoea?" *Health Policy Plan* 7(3), 260-268.
- Udry, Christopher. 1996. "Gender, Agricultural Production, and the Theory of the Household." *Journal of Political Economy*, 104(5), 1010-46.

- Vermeersch, Christel and Michael Kremer. 2005. "Schools Meals, Educational Achievement and School Competition: Evidence from a Randomized Evaluation," World Bank Policy Research Working Paper Series No.3523.
- World Bank. 2007. *Cost of Pollution in China: Economic Estimates of Physical Damages*. Washington, D.C.
- World Health Organization. 2004. "Water, Sanitation and Hygiene Links to Health, FACTS AND FIGURES – updated November 2004."  
[http://www.who.int/water\\_sanitation\\_health/publications/facts2004/en/](http://www.who.int/water_sanitation_health/publications/facts2004/en/) (accessed on Jan 3rd, 2013).
- World Health Organization. 2011. *World health statistics 2011*. France.
- Zhang, Jing. 2012. "The Impact of Water Quality on Health: Evidence from the Drinking Water Infrastructure Program in Rural China." *Journal of Health Economics*, 31(1), 122-34.
- Zhang, R.; H. Li; X. Wu; F. Fan; B. Sun; Z. Wang; Q. Zhang and Y. Tao. 2009. "Current Situation Analysis on China Rural Drinking Water Quality." *Journal of Environment and Health* (Chinese Version) 26 (1), 3–5.

**Table 1. Descriptive Statistics**

<b>Variables</b>	<b>Observations</b>	<b>Mean</b>	<b>Standard deviation</b>
Grades of education completed	4,729	8.696	(2.765)
Graduates of a middle school (yes/no)	4,729	0.658	(0.474)
Graduates of a high school (yes/no)	4,729	0.164	(0.370)
Age	4,729	23.202	(1.862)
Female	4,729	0.490	(0.500)
Household head's child (yes/no)	4,729	0.746	(0.435)
Household head's grandchild (yes/no)	4,729	0.051	(0.221)
Household size	4,729	5.132	(1.776)
Number of children in the household (age<=15)	4,729	0.167	(0.452)
Log of household income in the first wave	4,729	7.228	(1.444)
Distance to a primary school (Km)	4,729	0.571	(2.969)
Distance to a middle school (Km)	4,729	2.025	(5.332)
Distance to a high school (Km)	4,729	8.494	(11.564)

**Table 2. Regression Results for Education Indicators**

Dependent Variables	Grades of education completed (1)	Graduates of a middle school (yes/no) (2)	Graduates of a high school (yes/no) (3)
<b>Water plant</b>	1.081*** (0.178)	0.130*** (0.026)	0.147*** (0.024)
Age	0.121*** (0.022)	0.016*** (0.004)	0.012*** (0.003)
Female	-0.125 (0.094)	-0.036** (0.014)	0.011 (0.011)
Head's child	0.443*** (0.119)	0.009 (0.021)	0.038** (0.015)
Head's grandchild	0.242 (0.224)	-0.030 (0.042)	0.062** (0.029)
Household size	-0.073*** (0.028)	0.007 (0.005)	-0.016*** (0.004)
Number of children	-0.284** (0.113)	-0.066*** (0.021)	-0.010 (0.012)
Log of household income in the first wave	0.111*** (0.038)	0.012* (0.007)	0.007 (0.004)
Kms to a primary school	0.002 (0.009)	-0.004** (0.002)	0.000 (0.001)
Kms to a middle school	-0.024** (0.010)	-0.002 (0.002)	-0.002 (0.001)
Kms to a high school	-0.025*** (0.008)	-0.003** (0.001)	-0.002** (0.001)
Constant	3.398 (3.285)	0.166 (0.461)	-0.622 (0.465)
County-year FE	Yes	Yes	Yes
Observations	4,729	4,729	4,729
R-squared	0.255	0.194	0.180

Notes: Each column presents the results from separate regressions.

In addition to the covariates listed in the table, each regression also controls for county times year fixed effects.

The results in Column (2), (3), and (4) are from linear probability models.

The standard errors in parentheses are clustered at the village level.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

**Table 3. Regression Results for Older Adults**

Dependent Variables	Grades of education completed (1)	Graduates of a middle school (yes/no) (2)	Graduates of a high school (yes/no) (3)
<b>Water plant</b>	-0.031 (0.369)	0.021 (0.042)	-0.023 (0.043)
Age	-0.111*** (0.006)	-0.014*** (0.001)	-0.004*** (0.001)
Kms to a primary school	-0.028 (0.034)	-0.002 (0.004)	-0.001 (0.003)
Kms to a middle school	-0.054 (0.039)	-0.006 (0.004)	-0.007** (0.003)
Kms to a high school	-0.036*** (0.009)	-0.003*** (0.001)	-0.004*** (0.001)
Constant	0.646 (0.883)	-0.125 (0.093)	-0.148** (0.066)
Observations	2,708	2,708	2,708
R-squared	0.335	0.266	0.161

Notes: Each column presents the results from separate OLS regressions.

In addition to the covariates listed in the table, each regression also controls for county\*year fixed effects.

The standard errors in parentheses are clustered at the village level.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1.



**Table 4. OLS and IV estimation of Treatment Effects**

Dependent Variables	Grades of education completed		Graduates of a middle school (yes/no)		Graduates of a high school (yes/no)	
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Water plant</b>	1.020*** (0.186)	2.411*** (0.686)	0.128*** (0.027)	0.359*** (0.102)	0.130*** (0.024)	0.264*** (0.086)
Observations	4,246	4,246	4,246	4,246	4,246	4,246
R-squared	0.238	0.202	0.192	0.159	0.151	0.132
P-values (bootstrap Hausman test)	0.982		0.598		0.993	

Notes: Each column presents the results from separate regressions.

All of other covariates are the same with the ones in Table 2.

The results in Columns (1), (3), and (5) are from linear probability models, and the ones in columns (2), (4) and (6) are from 2SLS model when using the dummy of being non-flat as the instrument.

The standard errors in parentheses are clustered at the village level.

The bootstrap Hausman tests are based on 1000 bootstrap replications.

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

**Table 5. Correlation between the Treatment and Local Labor Market Conditions**

Dependent Variables	Daily male wage over average HH daily income in a village (1)	Daily female wage over average HH daily income in a village (2)	Daily construction workers' wage over average HH daily income in a village (3)	Daily childcare workers' wage over average HH daily income in a village (4)
<b>Water plant</b>	0.734 (0.538)	0.576 (0.417)	0.703 (0.702)	0.295 (0.274)
Observations	869	861	963	693
R-squared	0.497	0.607	0.489	0.694
<b>Controlling for the road conditions to capture time costs:</b>				
<b>Water plant</b>	0.791 (0.591)	0.602 (0.458)	0.658 (0.777)	0.454 (0.321)
Observations	843	834	931	669
R-squared	0.498	0.608	0.499	0.704

Notes: Each cell presents the results from separate regressions.

Each regression also controls for county\*year fixed effects.

In the survey road condition around the villages are described in three categories: paved, dirt and stone.

In the regressions, the two dummy variables—whether paved roads and whether stone roads around a village (default is the dirt roads)—are controlled in order to capture time costs needed for working outside of households.

The standard errors in parentheses are clustered at the village level.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

**Table 6. Treatment Effects with County-year and Village Fixed Effects**

Dependent variables	Grades of education completed		Graduates of a middle school (yes/no)		Graduates of a high school (yes/no)	
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Water plant</b>	1.081*** (0.178)	0.437** (0.214)	0.130*** (0.026)	0.050 (0.046)	0.147*** (0.024)	0.032 (0.028)
Fixed Effects	County-year	Village	County-year	Village	County-year	Village
Observations	4,729	4,729	4,729	4,729	4,729	4,729
R-squared	0.255	0.277	0.194	0.186	0.180	0.217

Notes: Each column presents the results from separate regressions.

All of other covariates are the same with the ones in Table 2.

The standard errors in parentheses are clustered at the village level.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

**Table 7. Treatment Effects with Controlling for Optimal Water Access**

Dependent variables	Grades of education completed		Graduates of a middle school (yes/no)		Graduates of a high school (yes/no)	
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Water plant</b>	1.014*** (0.173)	1.039*** (0.171)	0.116*** (0.025)	0.119*** (0.026)	0.145*** (0.024)	0.150*** (0.024)
Water access	0.741*** (0.212)		0.128*** (0.039)		0.034 (0.021)	
Observations	4,689	4,275	4,689	4,275	4,689	4,275
R-squared	0.259	0.248	0.198	0.191	0.181	0.187

Notes: Each column presents the results from separate regressions.

All of other covariates are the same with the ones in Table 2.

Column (1), (3) and (5) control for water access (=1 if the household has optimal water access; =0 if otherwise), and Column (2), (4) and (6) restrict the sample to observations with optimal water access (water access=1).

The standard errors in parentheses are clustered at the village level.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 8. Heterogeneous Treatment Effects across Income Groups**

Dependent variables	Grades of education completed (1)	Graduates of a middle school (yes/no) (2)	Graduates of a high school (yes/no) (3)
<b>Poor</b>			
<b>Water plant</b>	0.984*** (0.245)	0.107*** (0.039)	0.133*** (0.031)
Observations	2,381	2,381	2,381
R-squared	0.308	0.256	0.244
<b>Rich</b>			
<b>Water plant</b>	1.043*** (0.187)	0.128*** (0.027)	0.157*** (0.028)
Observations	2,348	2,348	2,348
R-squared	0.319	0.263	0.242

Notes: Each cell presents the results from separate regressions.

All of other covariates are the same with the ones in Table 2.

The standard errors in parentheses are clustered at the village level.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 9. Treatment effects over exposure time**

	Grades of education completed (1)	Grades of education completed (2)
<b>Water plant</b> (dummy)	0.652*** (0.156)	
Exposed to water plant: 0-2 years old		1.693*** (0.439)
Exposed to water plant: 3-5 years old		0.552 (0.395)
Exposed to water plant: 6-10 years old		0.524** (0.251)
Exposed to water plant: 11-15 years old		0.587** (0.226)
Exposed to water plant: 16-20 years old		0.447** (0.197)
Exposed to water plant: 21-25 years old		0.427** (0.165)
Other controls as in Table 2	Yes	Yes
Observations	4,064	4,064
R-squared	0.255	0.259

Notes: Each column presents the results from separate regressions.

All of other covariates are the same with the ones in Table 2.

The standard errors in parentheses are clustered at the village level.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 10. Effects on Education by Gender**

Dependent Variables	Boys			Girls		
	Grades of education completed (1)	Graduates of a middle school (2)	Graduates of a high school (3)	Grades of education completed (4)	Graduates of a middle school (5)	Graduates of a high school (6)
<b>Water plant</b>	0.874*** (0.197)	0.110*** (0.029)	0.138*** (0.029)	1.290*** (0.210)	0.151*** (0.032)	0.157*** (0.026)
Observations	2,414	2,414	2,414	2,315	2,315	2,315
R-squared	0.263	0.222	0.234	0.370	0.280	0.250

Notes: Each column presents the results from separate regressions.

All of other covariates are the same with the ones in Table 2.

The standard errors in parentheses are clustered at the village level.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table 11. Effects on Health by Gender**

Dependent variables	Boys		Girls	
	Body mass (i.e., weight/height) (1)	Height (2)	Body mass (i.e., weight/height) (3)	Height (4)
<b>Water plant</b>	0.545*** (0.170)	1.393*** (0.522)	0.386* (0.211)	1.086** (0.494)
Age	1.045*** (0.011)	3.580*** (0.037)	0.956*** (0.011)	2.938*** (0.041)
Household size	0.034 (0.048)	-0.554*** (0.146)	-0.022 (0.047)	-0.368** (0.166)
Log of household income in the first wave	0.125** (0.058)	0.245* (0.131)	0.044 (0.046)	0.030 (0.156)
Raising livestock	-0.257* (0.146)	-0.551 (0.403)	-0.118 (0.152)	-0.320 (0.423)
Distance to the nearest medical facility	0.086 (0.054)	0.649*** (0.153)	0.026 (0.057)	0.143 (0.165)
Constant	11.951*** (4.173)	82.046*** (7.461)	20.050*** (2.633)	113.419*** (9.444)
Observations	6,061	6,116	5,010	5,053
R-squared	0.827	0.872	0.815	0.828

Notes: Each column presents the results from separate regression.

The other control variables include county times year fixed effects.

The standard errors in parentheses are clustered at the village level.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1



**Table 12. Treatment Effects by Gender of the Elder Sibling**

Dependent Variables	With elder brother			With elder sister		
	Grades of education completed	Graduates of a middle school	Graduates of a high school	Grades of education completed	Graduates of a middle school	Graduates of a high school
<b>Water plant</b>	1.388*** (0.364)	0.179*** (0.062)	0.162*** (0.050)	0.571** (0.239)	0.058 (0.043)	0.103** (0.040)
Observations	844	844	844	839	839	839
R-squared	0.416	0.347	0.335	0.398	0.308	0.377

Notes: Each column presents the results from separate regressions.

All of other covariates are the same with the ones in Table 2.

The standard errors in parentheses are clustered at the village level.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table A1. Treatment Effects Based on Different Definitions of the Treatment Variable.**

Dependent Variables	Grades of education completed (1)	Graduates of a middle school (yes/no) (2)	Graduates of a high school (yes/no) (3)
	% of changes in plant water coverage per year between waves		
10%	1.120*** (0.167)	0.146*** (0.024)	0.147*** (0.023)
15%	1.042*** (0.172)	0.131*** (0.025)	0.136*** (0.023)
<b>Water plant</b> (20%)	1.081*** (0.178)	0.130*** (0.026)	0.147*** (0.024)
25%	1.063*** (0.181)	0.126*** (0.026)	0.144*** (0.024)
30%	1.151*** (0.190)	0.141*** (0.026)	0.155*** (0.026)
% of plant water coverage			
60%	1.156*** (0.165)	0.133*** (0.024)	0.167*** (0.023)
65%	1.113*** (0.168)	0.122*** (0.024)	0.166*** (0.023)
70%	1.135*** (0.165)	0.130*** (0.023)	0.164*** (0.023)
75%	1.109*** (0.177)	0.118*** (0.025)	0.158*** (0.025)
80%	1.127*** (0.165)	0.116*** (0.024)	0.165*** (0.026)
85%	1.178*** (0.169)	0.115*** (0.025)	0.171*** (0.027)

Notes: Each cell presents the results from separate regressions for different constructions of the treatment variable.

All of other covariates are the same with the ones in Table 2.

The standard errors in parentheses are clustered at the village level.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table A2. Treatment Effects with and without controlling for average village income and Gini coefficients**

Dependent variables	Grades of education completed		Graduates of a middle school (yes/no)		Graduates of a high school (yes/no)	
	baseline	With extra controls	baseline	With extra controls	baseline	With extra controls
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Water plant</b>	1.081*** (0.178)	0.964*** (0.162)	0.130*** (0.026)	0.117*** (0.025)	0.147*** (0.024)	0.133*** (0.023)
Observations	4,729	4,721	4,729	4,721	4,729	4,721
R-squared	0.255	0.266	0.194	0.199	0.180	0.190

Notes: Here the extra controls include average village income and Gini coefficient for every village.

Each column presents the results from separate regressions.

All of other covariates are the same with the ones in Table 2.

The standard errors in parentheses are clustered at the village level.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

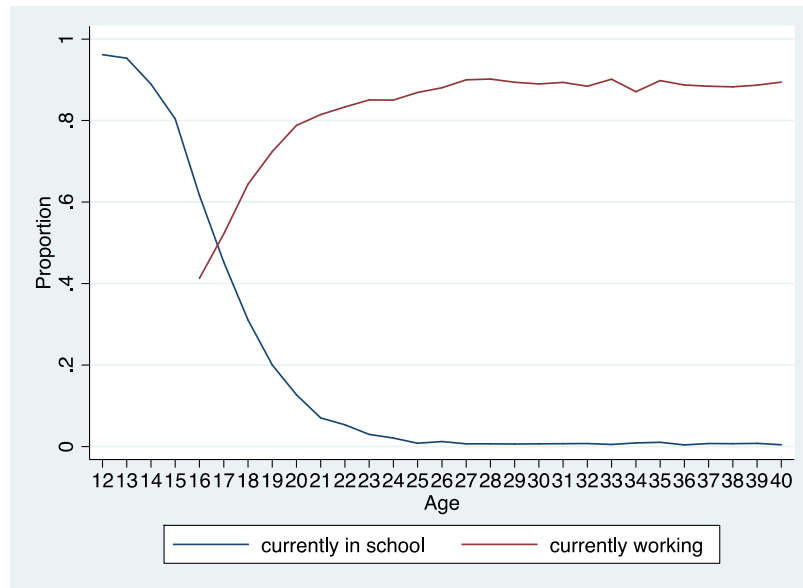
**Table A3. Mean Differences between Characteristics of Treated and Untreated Villages (at village level)**

<b>Variables</b>	<b>observations</b>	<b>Mean</b>	<b>Standard error</b>
Age	591	-0.264	(0.204)
Female	591	0.033	(0.071)
Household head's child (yes/no)	591	0.048	(0.062)
Household head's grandchild (yes/no)	591	-0.003	(0.030)
Household size	591	0.239	(0.262)
Number of children in the household (age<=15)	591	-0.057	(0.075)
Log of household income	576	0.232	(0.236)
Distance to a primary school (Km)	591	-0.050	(1.044)
Distance to a middle school (Km)	591	0.193	(1.022)
Distance to a high school (Km)	591	-2.031	(3.366)

Notes: the means of the treated villages are the average of their characteristics in five years before the treatment. The mean differences are adjusted for county-year fixed effects and the standard errors in parentheses.

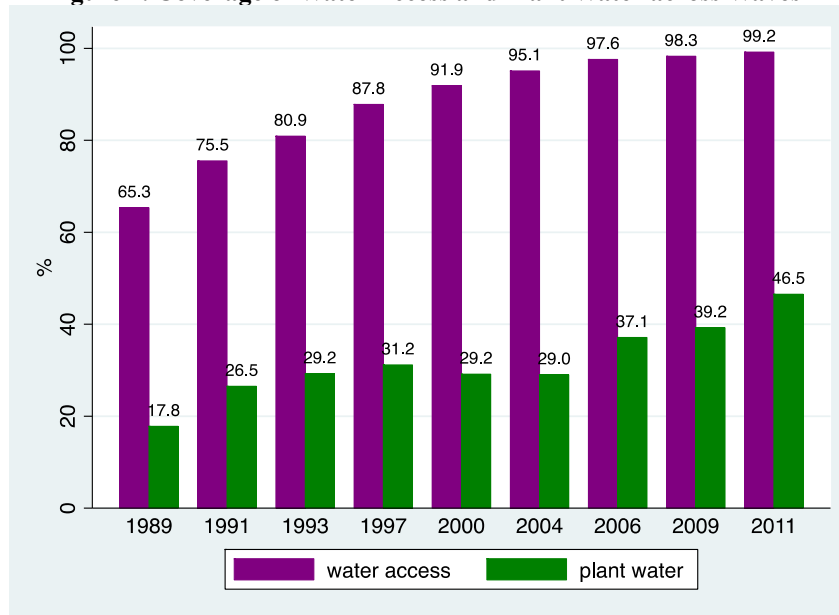
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

**Figure 1. Educational and Work Status of Each Age Group in Rural China**



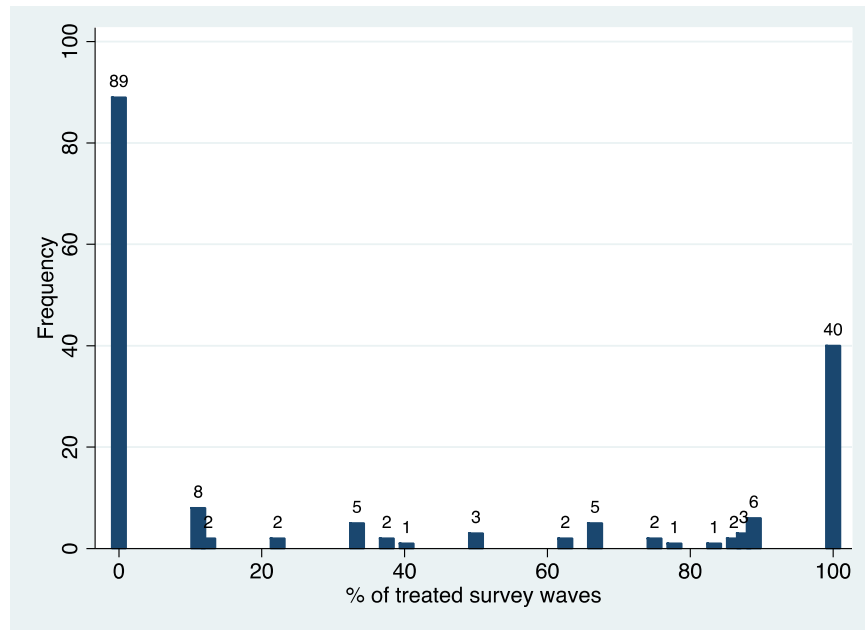
Data source: China Health and Nutrition Survey (CHNS)

**Figure 2. Coverage of Water Access and Plant Water across Waves**



Data Source: China Health and Nutrition Survey (CHNS)

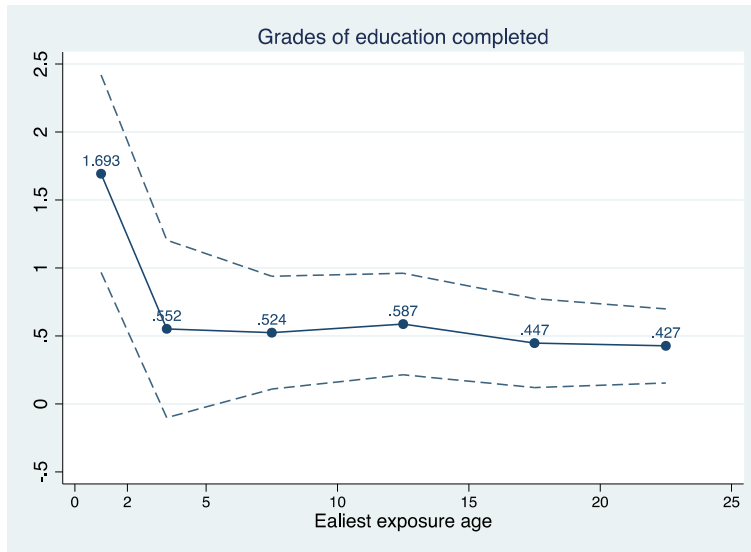
**Figure 3. Distribution of % of treated survey waves**



Data Source: China Health and Nutrition Survey (CHNS)

Notes: since villages may show up in different amount of periods, we calculate their percent of treated survey waves by dividing the number of treated waves by the number of waves when they showed up in the survey. And the vertical axis presents the number of villages who have the same amount of percent of treated survey waves.

**Figure 4. Educational Gain over earliest exposure age**



Data Source: China Health and Nutrition Survey (CHNS)

Notes: The estimated coefficients of the dummy variables indicating a child's earliest exposure age and their 95 percent confidence intervals (in dashed lines).